Supplementary Materials

Table of content

Supplementary Text

Text S1- Influence of admixture and sample size on genic SD SNPs proportions in the ASW population

In order to explore the possible effect of recent (non sex-specific) admixture on genic SD SNPs proportions, we simulated ten admixed populations (ADP_{1-10}) by randomly sampling individuals half from CEU and half from YRI and merging them into new populations with the same sample size and sex-ratio as ASW. The source populations, CEU and YRI, have been chosen because of the known admixed history of the ASW population. Because one of the source populations of the MEX population was not available here (i.e., Amerindian populations), we did not perform this analysis on this population.

ASW population has also a much smaller sample size than CEU and YRI populations; therefore the exact Fisher's test is less powerful to detect genetic differentiation between the sexes in this population. To discriminate the admixture effect from a potential sample size bias, we therefore also compared the proportions of genic SD SNPs between ASW and ten subsamples of CEU ($CEUR_{1-10}$) and ten subsamples of YRI ($YRIR_{1-10}$) with the same number of individuals than ASW and the same sex ratio.

We computed the 99.9% confidence intervals of the mean proportions of genic SD SNPs over the simulated admixed populations ADP_{1-10} and over the subsamples $CEUR_{1-10}$ and $YRIR_{1-10}$. We then aimed to compare the proportions observed in the original HapMap populations and those observed in the simulated populations.

For the X chromosome, the proportion of genic SD SNPs in ASW is not included in the confidence interval of neither the $CEUR_{1-10}$ and $YRIR_{1-10}$ populations, nor the ADP_{1-10} populations (Figure S7A-B). On the contrary, the proportions of genic SD SNPs in ASW on autosomes are always included in at least one of the $CEUR_{1-10}$ or $YRIR_{1-10}$ confidence intervals, except for chromosome 3, and most of them are included in the ADP_{1-10} confidence intervals. Therefore, the mean proportions of genic SD SNP over the resized populations and the simulated admixed populations are overall equal to the autosomal genic SD SNP proportions observed in ASW, while they are significantly higher than the X chromosome genic SD SNP proportion observed in ASW.

These results suggest that, for the autosomes, the lower proportion of genic SD SNPs observed in ASW compared to CEU and YRI could be due to a sample size bias. On the contrary, the lower proportion of genic SD SNPs in ASW observed for the X chromosome cannot be explained by a sample size and/or recent admixture effect. Moreover, it is interesting to see that the effect of admixture combined to sample size and the effect of sample size alone on the proportions of genic SD SNP are the same; therefore it is difficult to discriminate the effect of admixture.

Text S2- Functional analysis of the SD SNPs with extreme signals of F_{ST} after correction for multiple testing at the genome-wide level

To further investigate the functional significance of genome-wide SD SNPs, we searched for potential regulatory regions in their vicinity using the UCSC genome browser [1], also considering neighboring SNPs in high Linkage Disequilibrium (LD).

The linkage disequilibrium (LD) was calculated as a r^2 value between every possible pair of SNPs from each set including the extreme SD SNP and ± 50 SNPs using the formula from the R package genetics version 1.3.8 [2] re-implemented in Python.

Both X-linked SNPs identified are in high LD with each other ($r^2 > 0.8$), but are isolated from their neighbouring SNPs which do not show any sex-specific differences in allelic frequencies (Figure S14A-B). These patterns are consistent with the expectation that loci undergoing SA selection should be in LD with each other [3].

We examined the chromatin state segmentation based on epigenetics markers (Broad ChromHMM, integrating ChIP-seq data for eight common histone modifications) which indicates whether the region might be transcribed or not. We also looked at the Transcription Factor Binding Sites (TFBS conserved sites track) identified by ChIP-seq assays and the DNaseI hypersensitivity sites (ENCODE DNase Cluster track). The DNaseI cuts DNA in unoccupied regions; therefore a region of DNaseI hypersensitivity is likely to be a binding site of a transcription factor. Finally, we looked for signals of Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE, UNC FAIRE track) assays that determine if the region is associated with a regulatory activity by revealing which regions are depleted of nucleosome. We also looked for potential eQTLs in the region on the University of Chicago QTL Browser (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/).

We found no signal of FAIRE or eQTLs in any of the regions studied.

For the chromosomal SD SNPs showing extreme signals of SA polymorphisms on the X chromosome, rs5943145 maps in a repressed region for one cell type and is located in an intron of DCX, whose protein product directs neuronal migration. Interestingly, defects in this gene cause different diseases between the sexes: a "double cortex" syndrome in females and a "smooth brain" syndrome in males, suggesting that the protein has a different role in both sexes. The extreme SD SNP rs7053144 maps in an intron of ZCCHC16, a zinc finger protein.

References

1. Karolchik, D., Kuhn, R.M., Baertsch, R., Barber, G.P., Clawson, H., Diekhans, M., Giardine, B., Harte, R.A., Hinrichs, A.S., Hsu, F., et al. (2008). The UCSC Genome Browser Database: 2008 update. Nucleic Acids Res. *36*, D773–D779.

2. Warnes, G., Gorjanc, G., Leisch, F., and Man, M. (2013). genetics: Population Genetics.

3. Ubeda, F., Haig, D., and Patten, M.M. (2011). Stable linkage disequilibrium owing to sexual antagonism. Proc. R. Soc. B *278*, 855–862.

Supplementary Figures

Figure S1- Effect of FDR corrections at the gene level in assessing SD SNPs. P-values of Spearman's rank correlation tests between the length of a gene and its number of SD SNPs **(A)** before and **(B)** after FDR correction at the gene level. P-values lower than 10^{-5} are assigned to the darkest blue and p-values greater than 0.05 are shown in white.

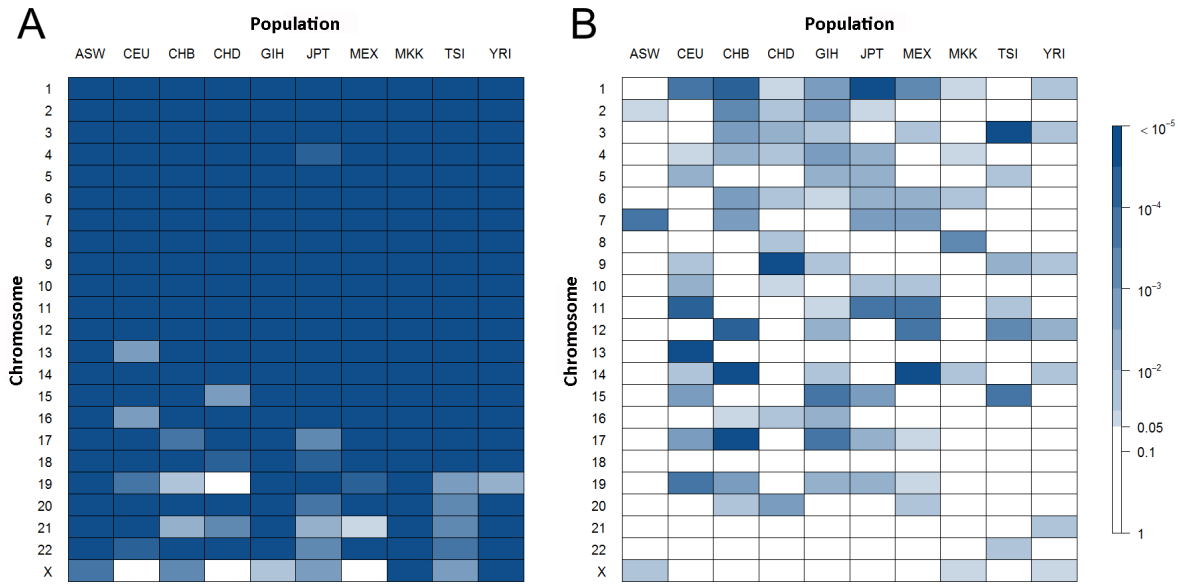


Figure S2- Method for generating bipartitions of individuals in the HapMap populations with different sex-ratio bias (Δ_p). The pipeline is shown for autosomes and the X chromosome using the CEU population's sample sizes. $\Delta_p = | p_{fA} - p_{fB} |$, where p_{fA} is the proportion of females in group A and p_{fB} the proportion of females in group B.

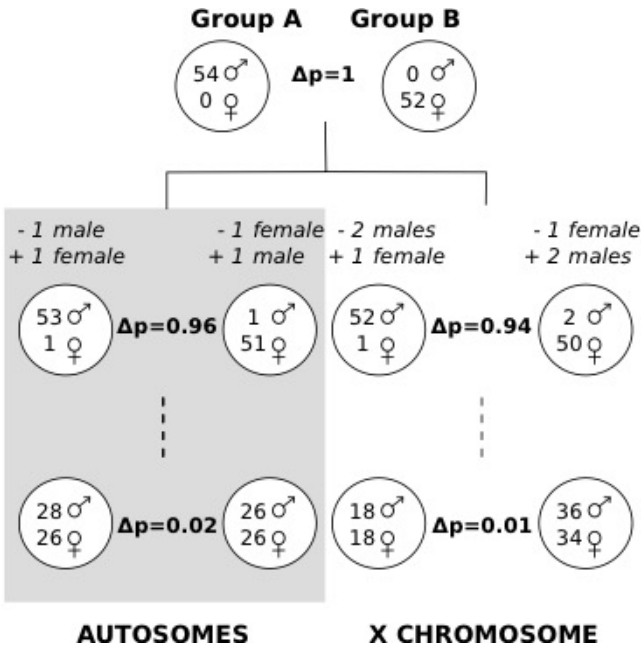
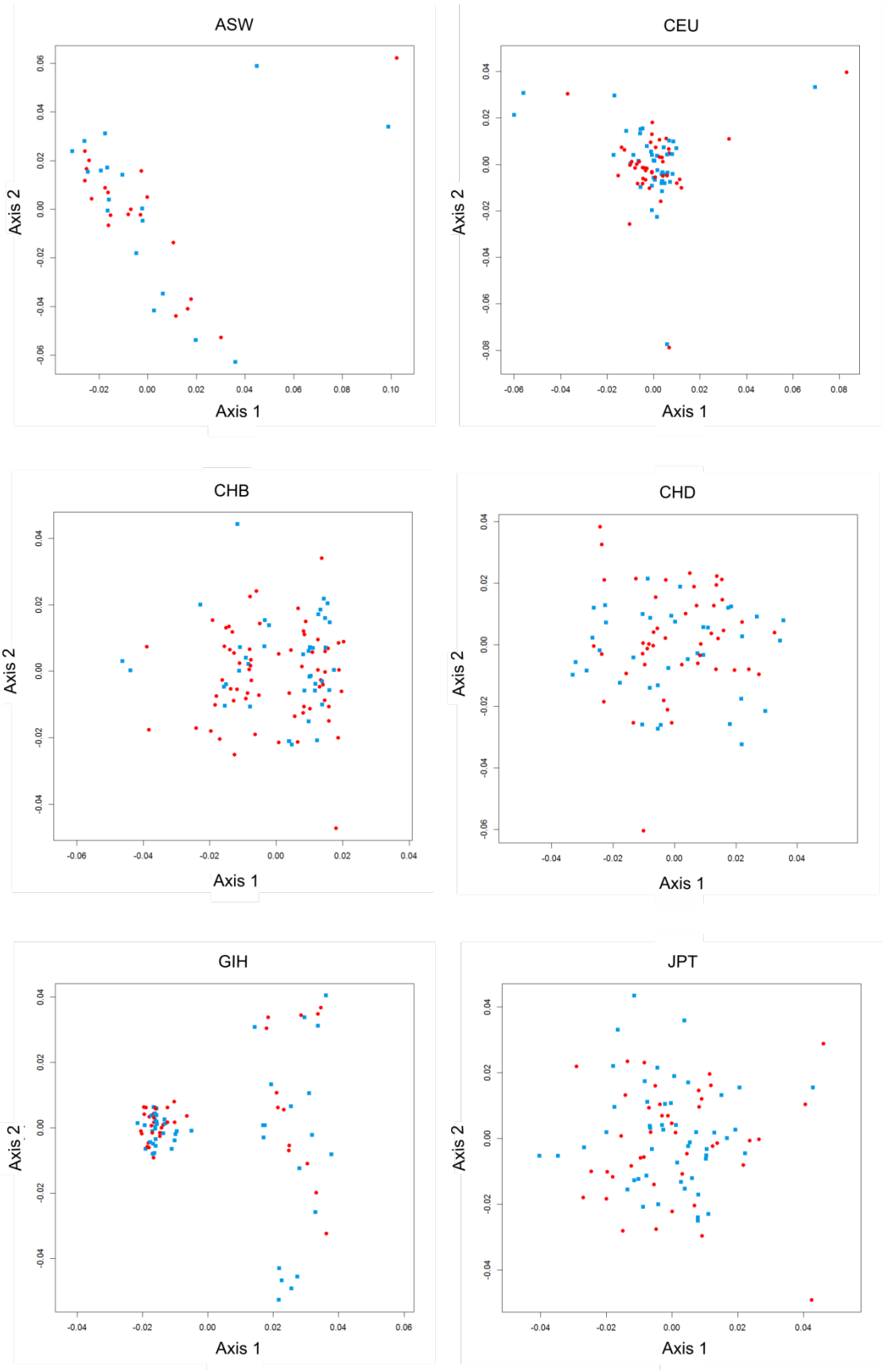
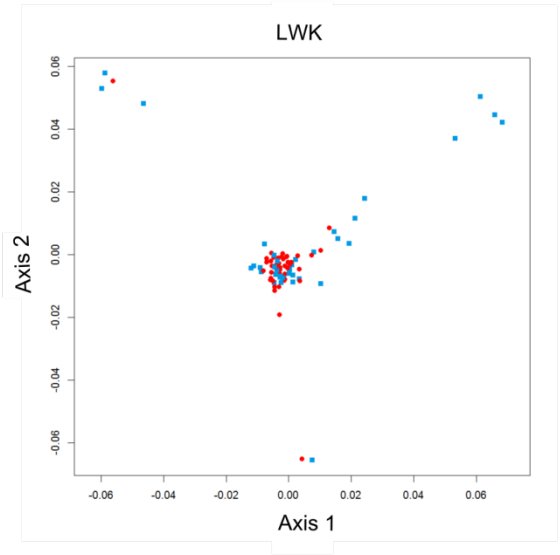
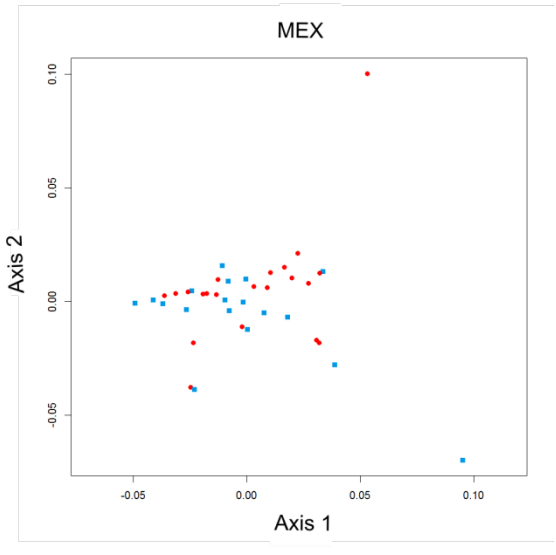
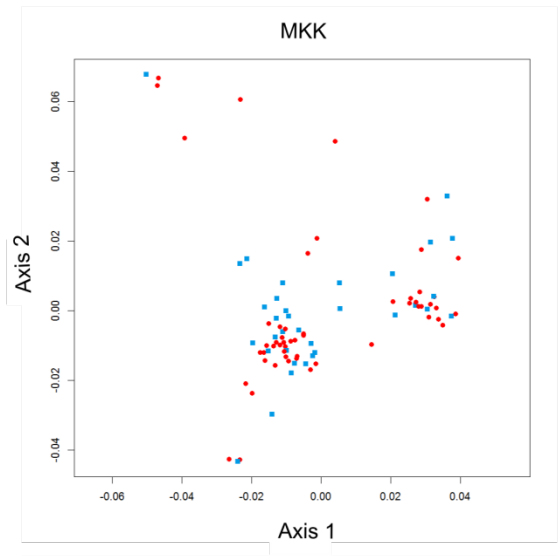


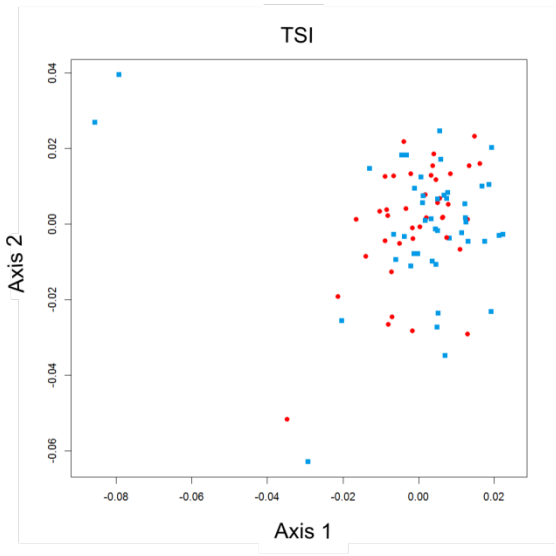
Figure S3- MDS plots computed from autosomal genetic data. Blue squares: male individuals, red dots: female individuals.




LWK
Axis 2
Axis 1
-0.06
-0.04
-0.02
0.00
0.02
0.04
0.06


MEX
Axis 2
Axis 1
-0.05
0.00
0.05
0.10


MKK
Axis 2
Axis 1
-0.06
-0.04
-0.02
0.00
0.02
0.04
0.06


TSI
Axis 2
Axis 1
-0.08
-0.06
-0.04
-0.02
0.00
0.02
0.04

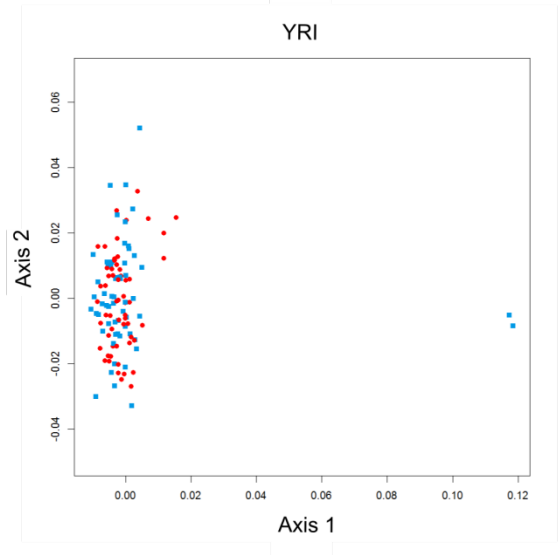
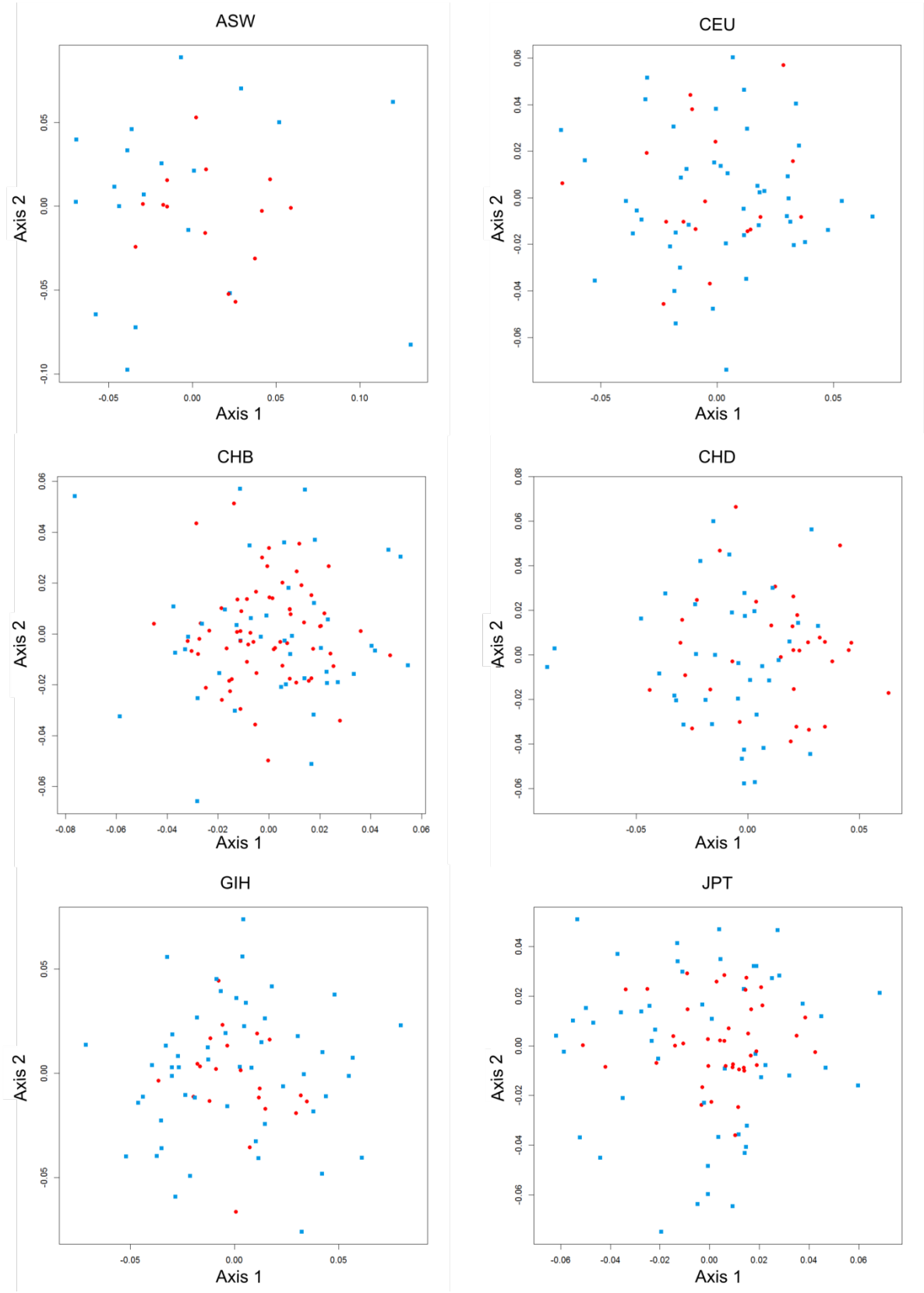

YRI
Axis 2
Axis 1
-0.04
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12

Figure S4- MDS plots computed from X-linked genetic data for the first female subsampling (out of 30). Blue dots: male individuals, red dots: female individuals.



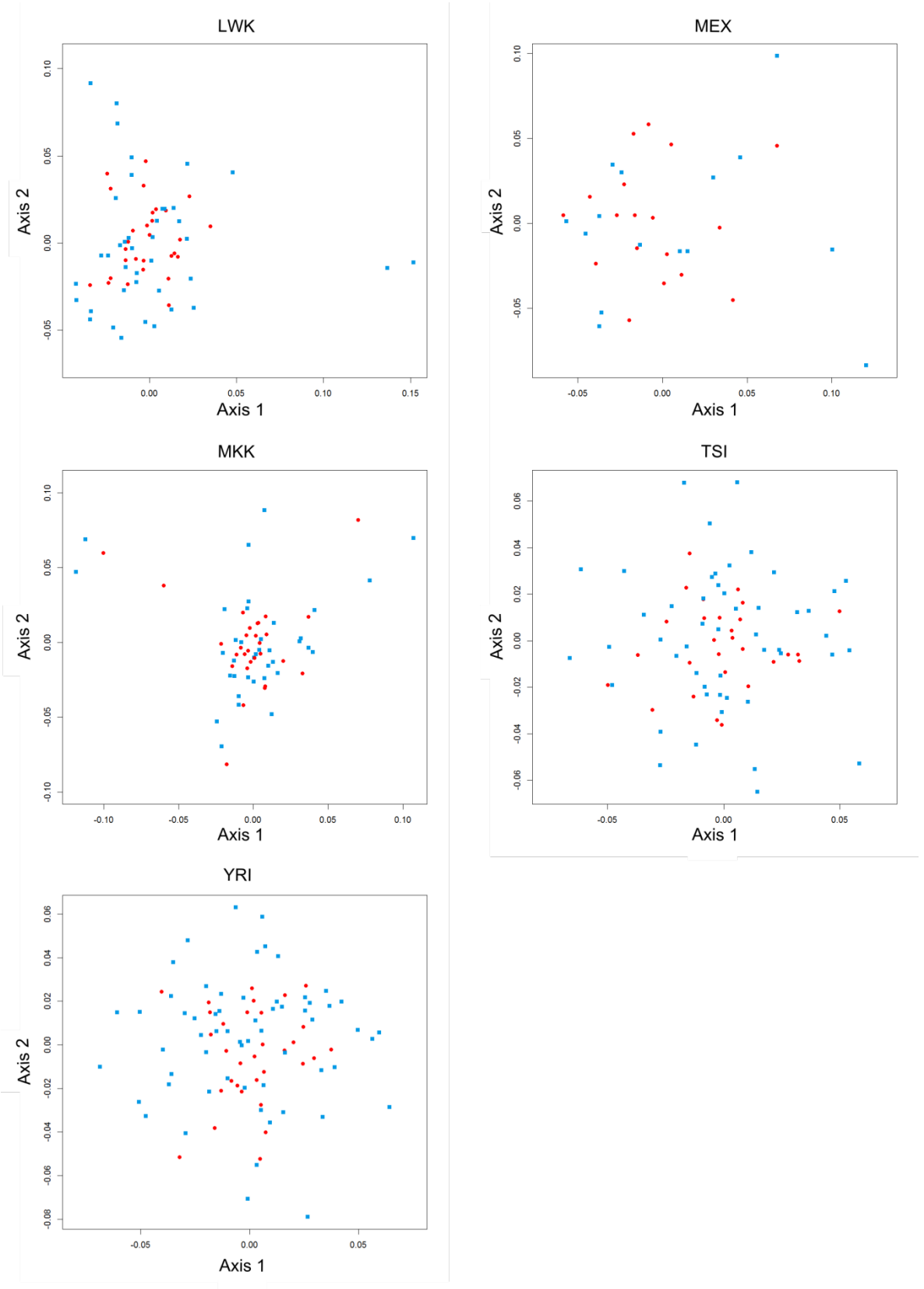

LWK
Axis 2
Axis 1
MEX
Axis 2
Axis 1
MKK
Axis 2
Axis 1
TSI
Axis 2
Axis 1
YRI
Axis 2
Axis 1

Figure S5- Effect of haplodisation of autosomes in males on the mean proportion of genic SD SNPs per chromosome. The distribution of the proportions of genic SD SNPs after FDR correction at the gene level over the 10 HapMap populations is shown per chromosome. Levels of significance of one-sided Wilcoxon–Mann–Whitney tests between the mean proportions of the X chromosome and each of the autosomes are reported (p-value: *< 0.05, **< 10^{-3}, ***< 10^{-4}).

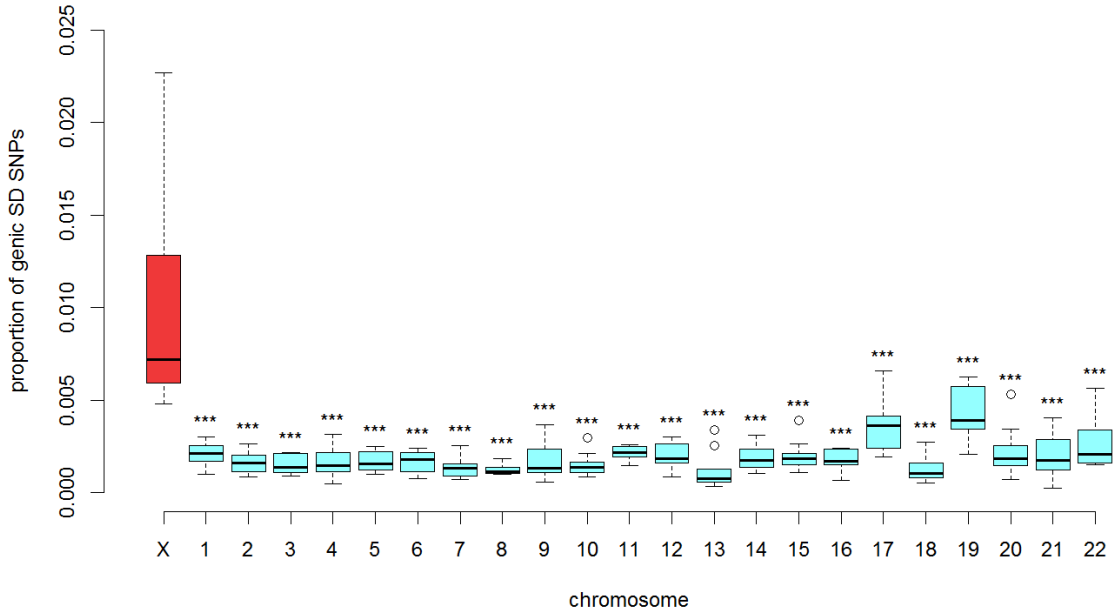


Figure S6- Concordance between the HapMap and 1000 Genomes datasets. Only genic SD SNPs showing similar frequency patterns between the two datasets are included in the analysis. The distribution of proportions of genic SD SNPs after FDR correction over the 10 HapMap populations is shown per chromosome. Levels of significance of one-sided Wilcoxon–Mann–Whitney tests between the mean proportions of the X chromosome and each of the autosomes are reported (p-value: *< 0.05, **< 10^{-3}, ***< 10^{-4}).

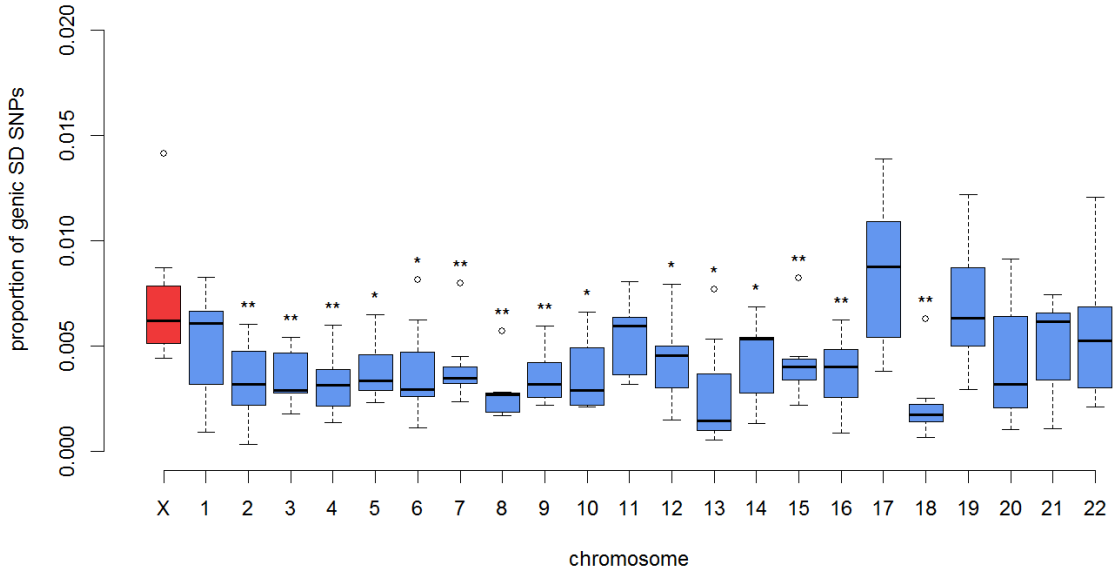


Figure S7- Influence of admixture and sample size on genic SD SNPs proportions.
(A) Simulation of admixture: For each chromosome, mean proportions of genic SD SNPs per chromosome (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) over 10 simulated admixed populations (ADP) and proportion of genic SD SNPs in ASW, YRI and CEU populations after FDR correction at the gene level. The error bars represent the 99.9% confidence intervals of the expected genic SD SNPs proportions in simulated admixed populations.
(B) Resized populations: For each chromosome, mean proportions of genic SD SNPs per chromosome (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) in simulated resized populations (10 CEUR and 10 YRIR) and proportion of genic SD SNPs in ASW, YRI and CEU, after FDR correction at the gene level. Error bars represent the 99.9% confidence intervals of the expected genic SD SNPs proportions in resized populations.

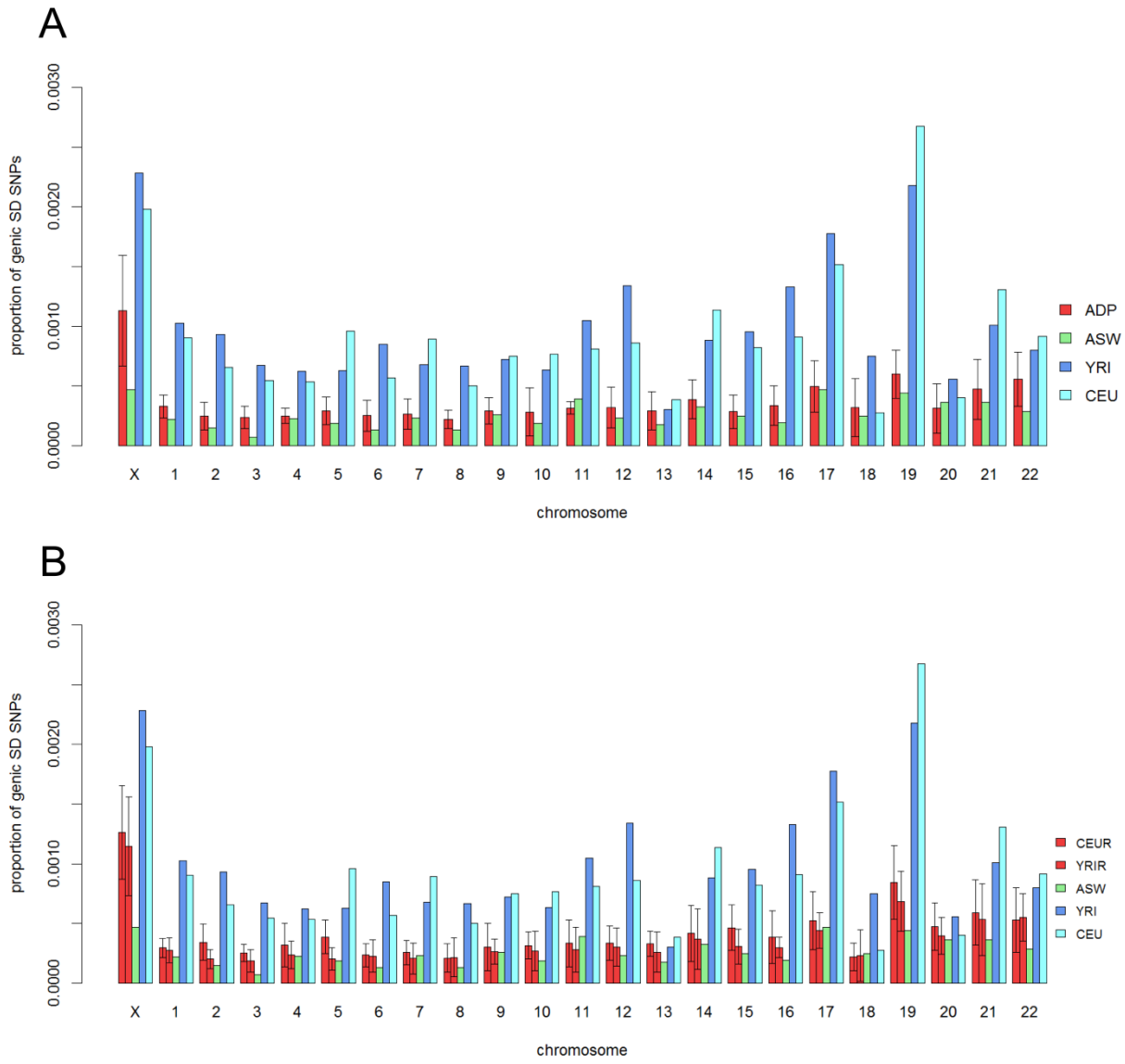


Figure S8- Proportions of genic SD SNPs per chromosome after LD pruning (filtering of genic SNPs with a $r^2 > 0.1$) **without the admixed populations** ASW and MEX. For each chromosome, the distributions of these proportions after FDR correction over the 10 HapMap populations are represented by boxplots. A one-sided Wilcoxon–Mann–Whitney test was performed to compare these distributions between the chromosome X (in red) and each autosome (in blue) (p-value: *< 0.05, **< 10^{-2}, ***< 10^{-3}).

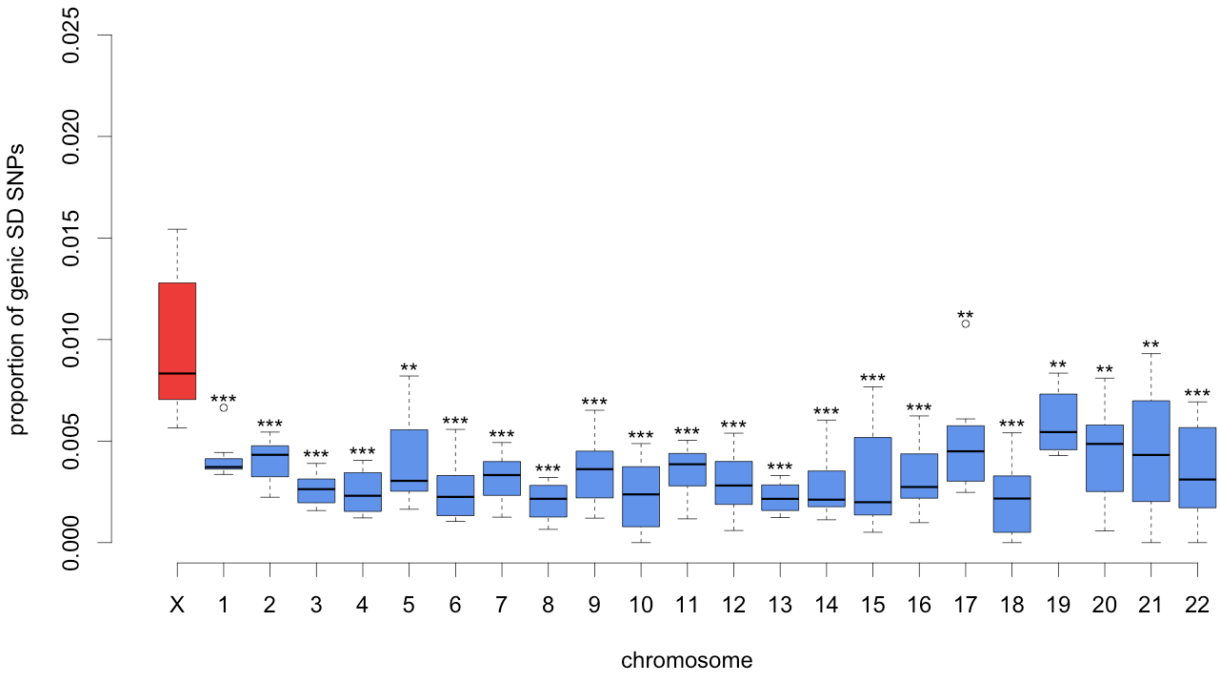
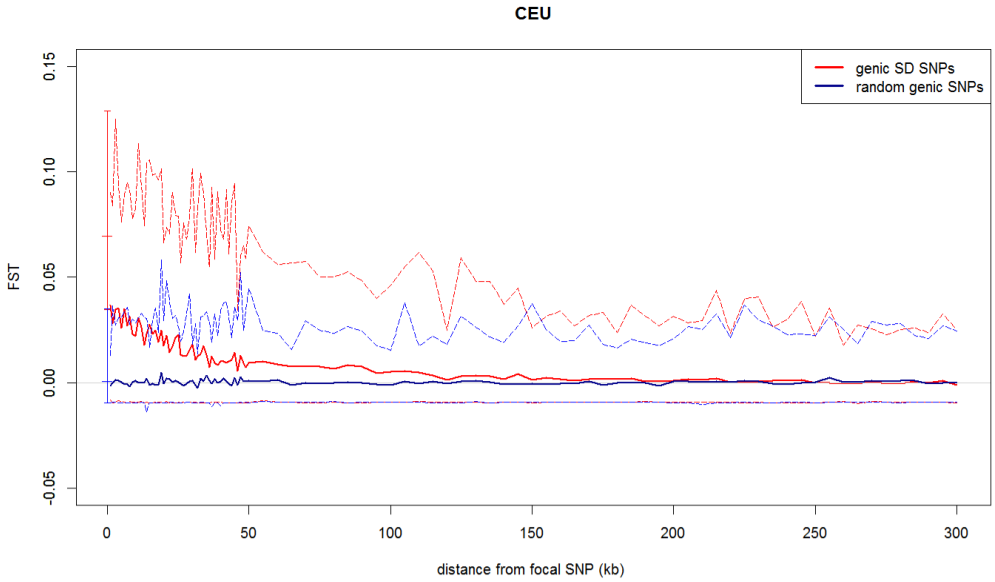
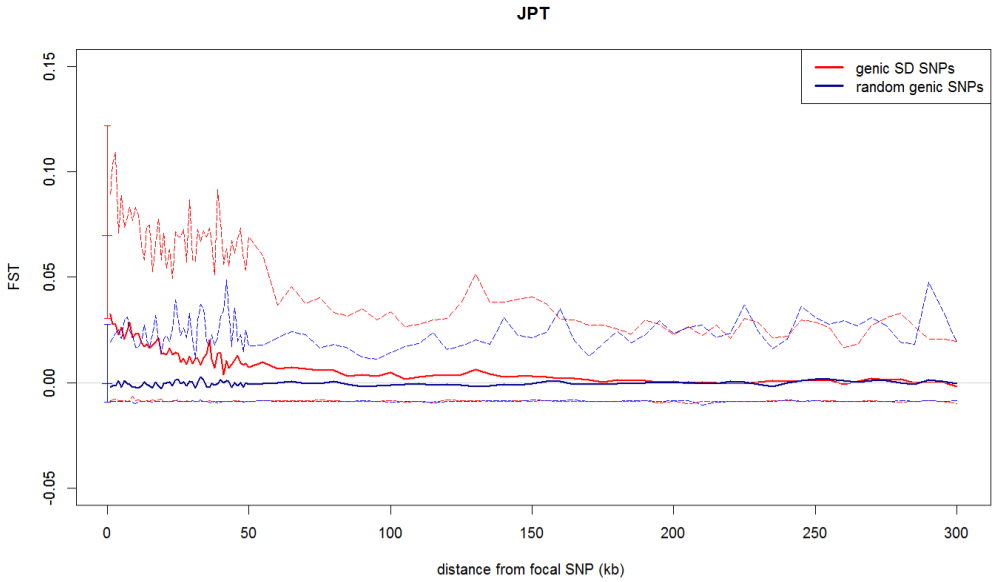


Figure S9- F_{ST} distribution in region neighboring the genic SD SNPs after LD pruning for the CEU, JPT and YRI populations. The F_{ST} were computed per SNPs, and then averaged per interval. We considered intervals of 1 kb for the first 50 kb, and then intervals of 5 kb between 50 kb and 300 kb. The mean F_{ST} around genic SD SNPs are shown in red, the mean F_{ST} around random genic SNPs are shown in blue. The dotted lines represent the 95% quantiles of the mean F_{ST} per interval.


CEU
genic SD SNPs
random genic SNPs
FST
0.15
0.10
0.05
0.00
-0.05
0
50
100
150
200
250
300
distance from focal SNP (kb)


JPT
genic SD SNPs
random genic SNPs
FST
0.15
0.10
0.05
0.00
-0.05
0
50
100
150
200
250
300
distance from focal SNP (kb)

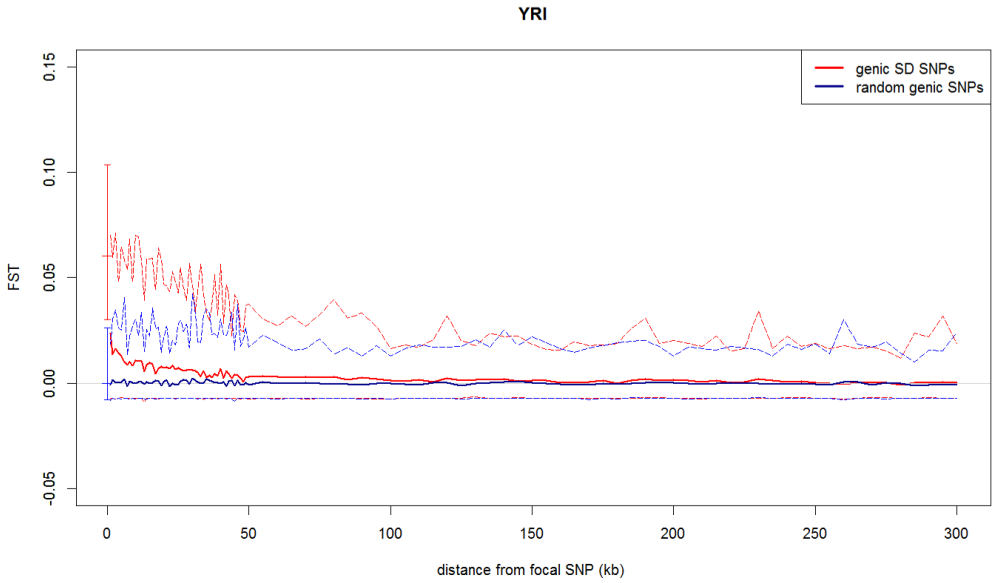

YRI
genic SD SNPs
random genic SNPs
FST
0.15
0.10
0.05
0.00
-0.05
0
50
100
150
200
250
300
distance from focal SNP (kb)

Figure S10- Significance of the F_{ST} signal around genic SD SNPs as compared to random genic SNPs. A one-sided Wilcoxon–Mann–Whitney test was performed for each interval to assess whether the F_{ST} signal in the vicinity of genic SD SNPs was higher than the one observed around random genic SNPs. When significant ($p < 0.05$), the corresponding interval is depicted in blue. From 0 to 50kb, the interval width is 1kb, while it is 5kb from 50 to 300kb. From 0kb until 50kb, the step between each interval is of 1kb, while from 50kb until 300kb, the step is of 5kb. The intervals before 1kb are not shown because the p-values of the Wilcoxon–Mann–Whitney are significant for each population.

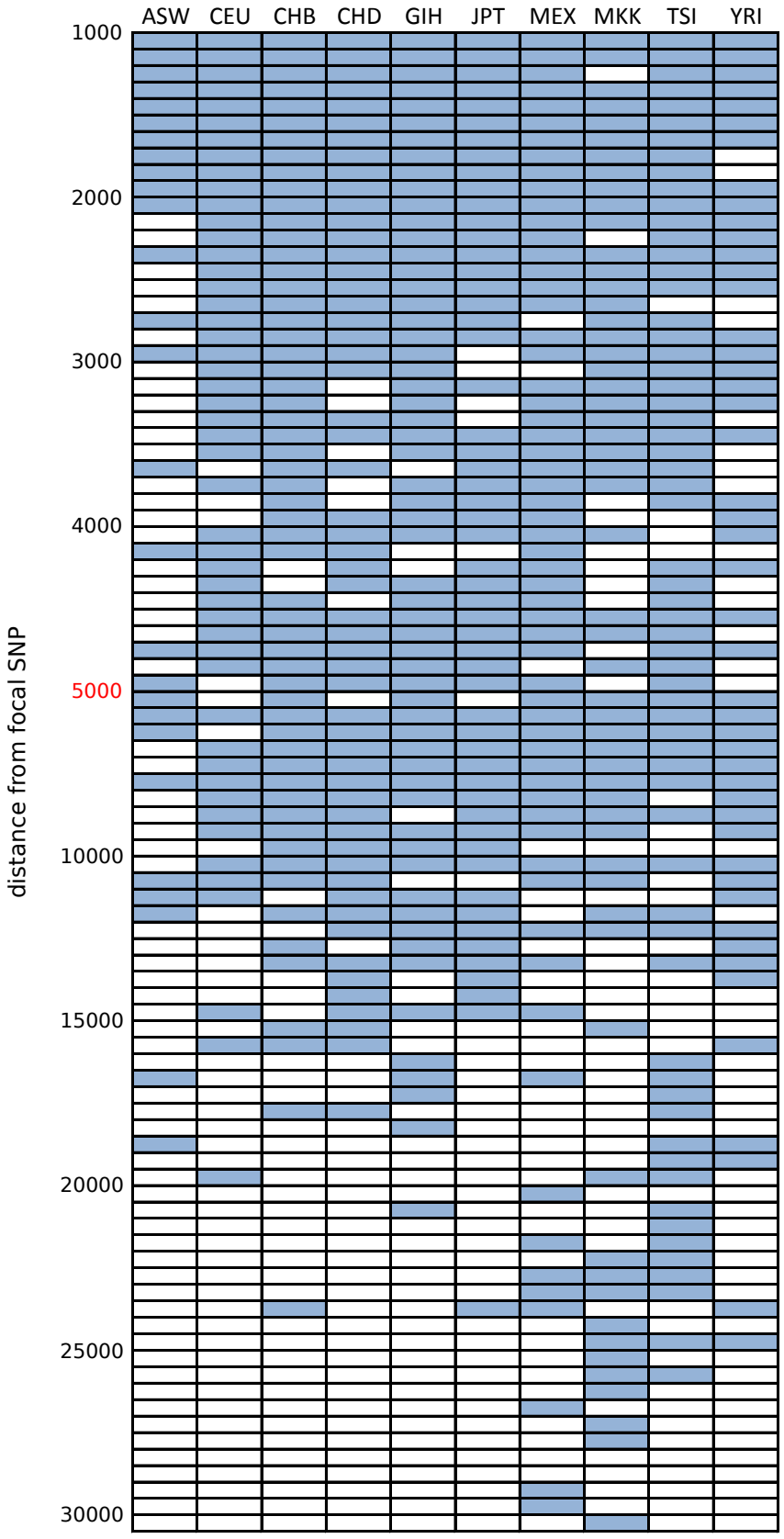

ASW
CEU
CHB
CHD
GIH
JPT
MEX
MKK
TSI
YRI
1000
2000
3000
4000
5000
10000
15000
20000
25000
30000
distance from focal SNP

Figure S11- Percentage of similarity between the set of genic SD SNP detected when comparing groups composed exclusively of males or females and the sets with different sex-ratio bias Δ_p, after LD-pruning, in blue for the autosomes and in red for the X chromosome. The percentage of similarity between the 10 repetitions within each Δ_p is indicated in blue for the autosomes and in red for the X chromosome.

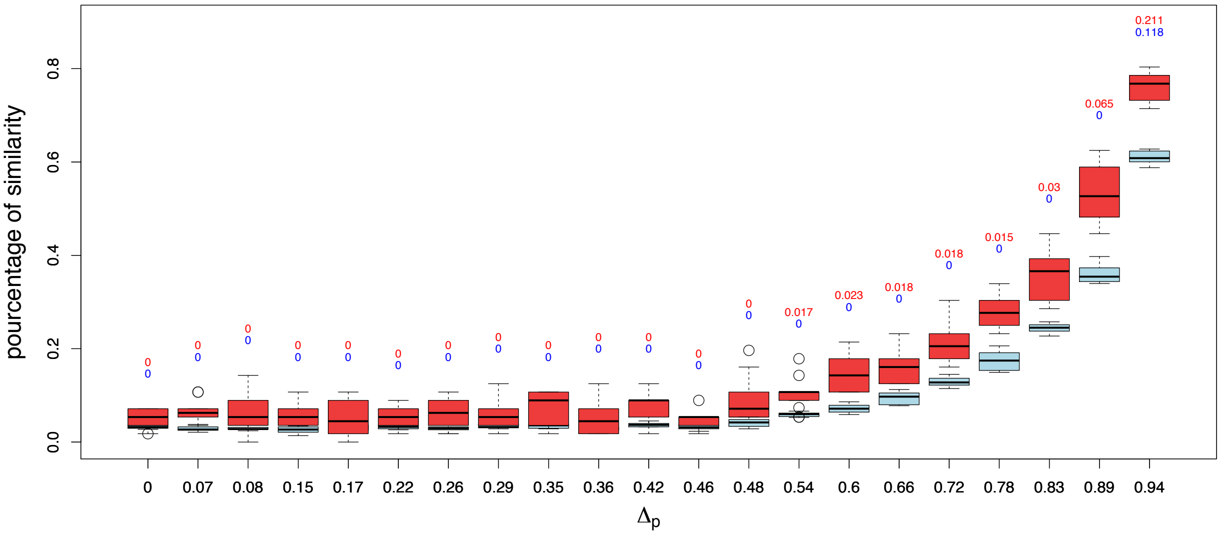


Figure S12- F_{ST} distributions of a simulated X-linked locus under SA selection for different selection and dominance coefficients A- for 3,000 individuals, B- after 10 resampling of 100 individuals. s_f is the selection coefficient in females ($s_f = s_m$), and h_f is the dominance coefficient in females ($h_f = 1-h_m$). The red dotted line corresponds to the mean F_{ST} observed in the non-admixed HapMap populations for the X-linked genic SD SNPs.

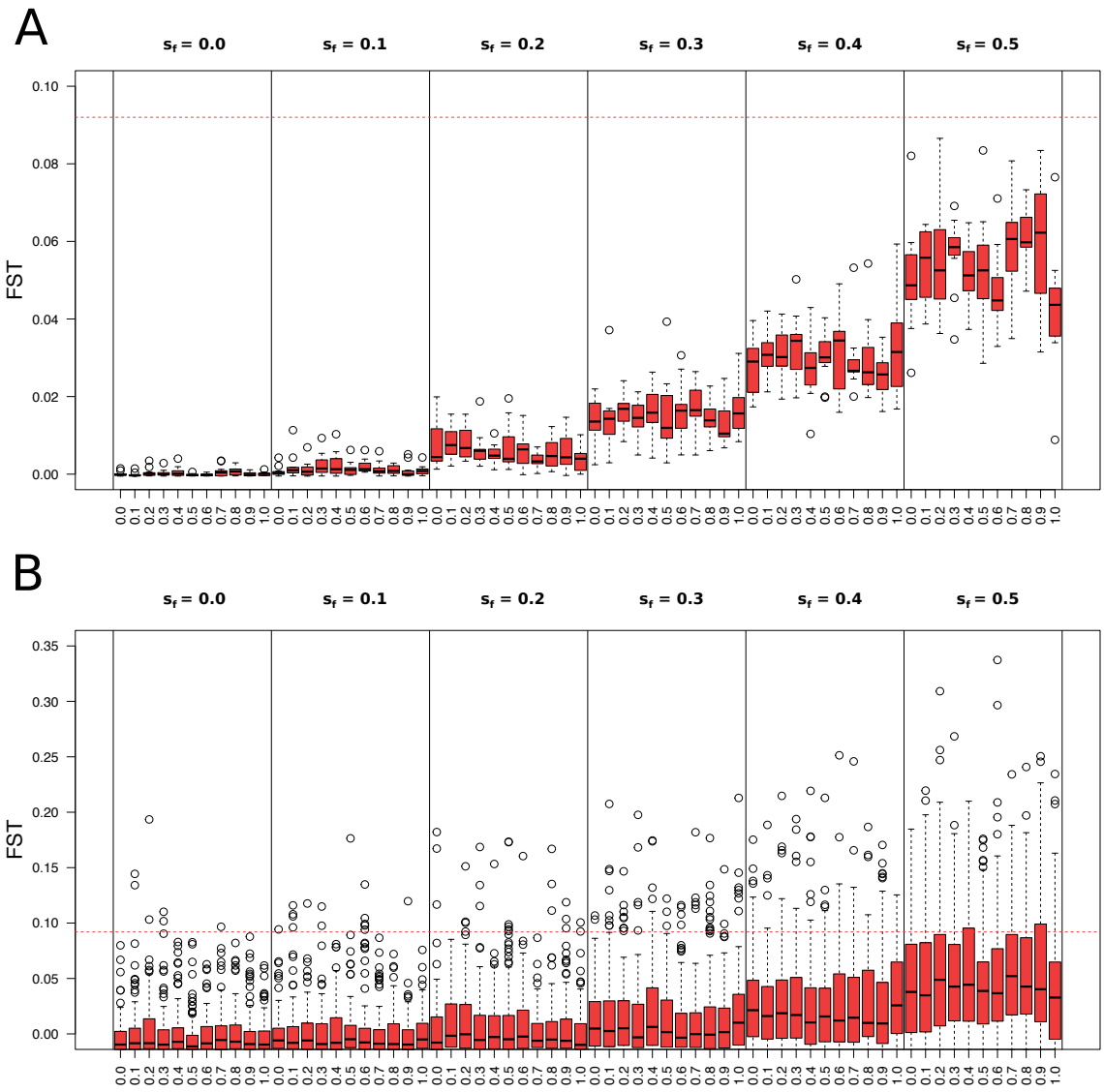


Figure S13- F_{ST} distributions of a simulated autosomal locus under SA selection for different selection and dominance coefficients A- for 3,000 individuals, B- after 10 resampling of 100 individuals. s_f is the selection coefficient in females ($s_f = s_m$), and h_f is the dominance coefficient in females ($h_f = 1\text{-}h_m$). The blue dotted line corresponds to the mean F_{ST} observed in the non-admixed HapMap populations for the autosomal genic SD SNPs.

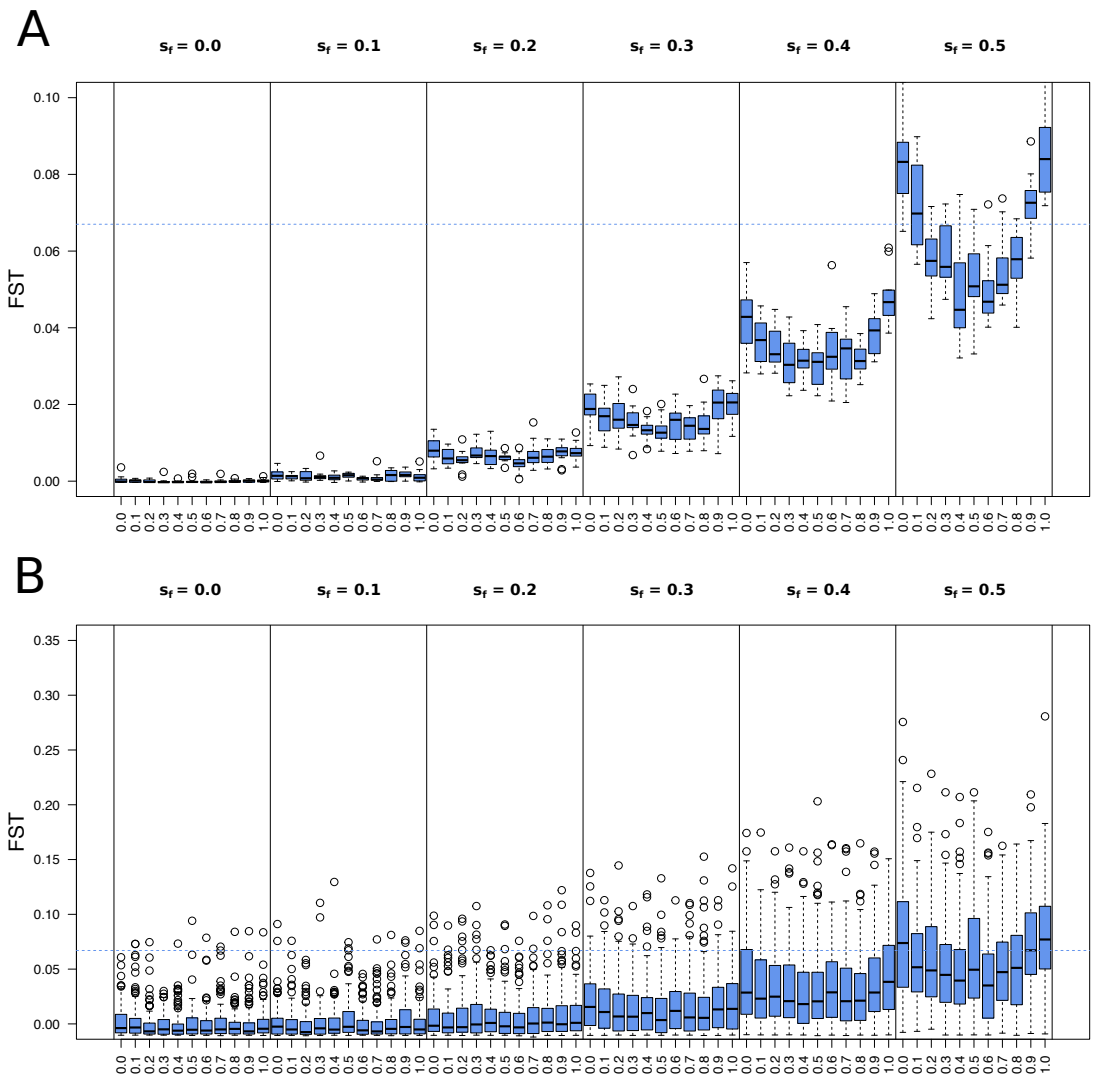
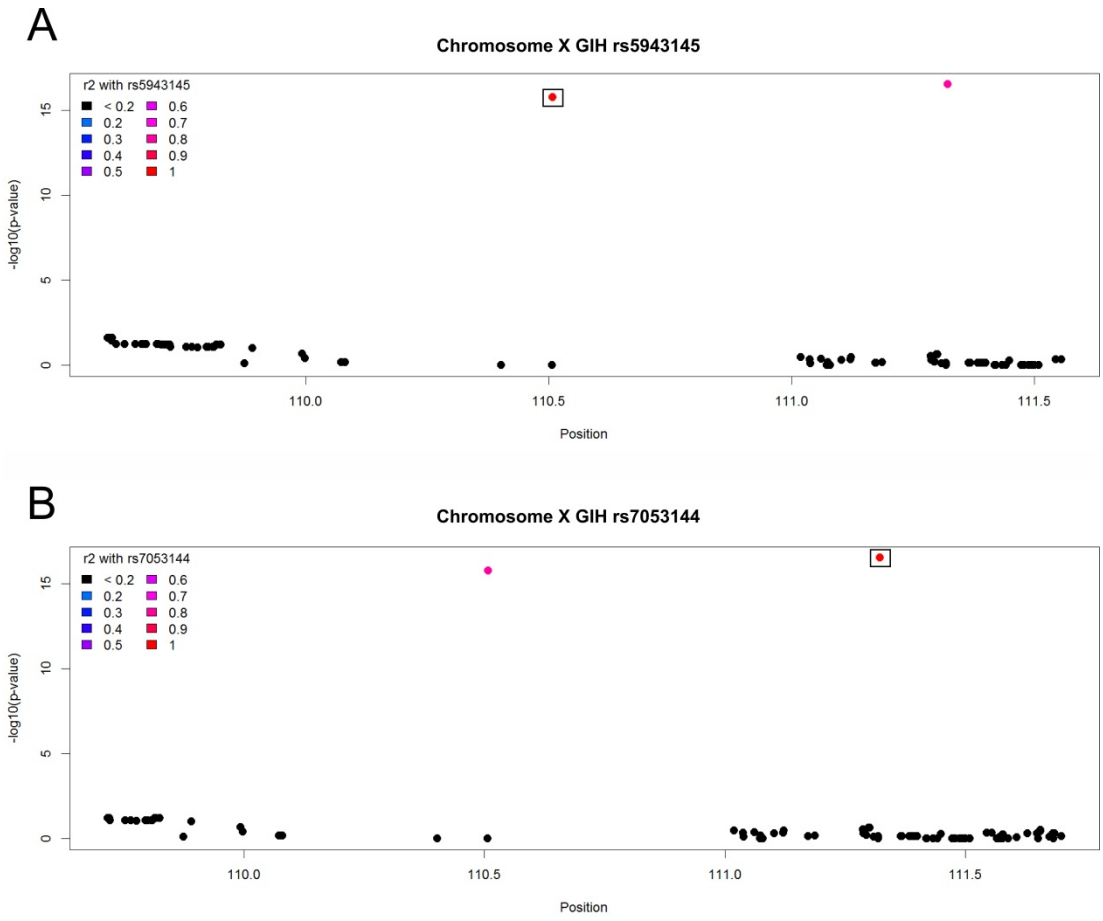


Figure S14- Analysis of the neighboring SNPs of the SD SNPs found on chromosome X with extreme signals of F_{ST} after correction for multiple testing at the genome-wide level. P-values of the F_{ST} between males and females for SNPs in the neighborhood of focal SD SNPs (in the boxes): **(A)** rs5943145 (chromosome X – GIH population), **(B)** rs7053144 (chromosome X – GIH population). Colors (from dark blue to red) indicate increasing LD (as quantified by r2) with the focal SNP. Positions were extracted from build NCBI36/hg18.



Supplementary Tables

Table S1. Number of individuals in the HapMap HAP1161 dataset in each population after quality control.

Population		Number of individuals		
		Total	Male	Female
ASW	African ancestry in Southwest USA	41	19	22
CEU	Utah residents with Northern and Western Europe ancestry	106	54	52
CHB	Han Chinese in Beijing, China	137	53	84
CHD	Chinese in Denver, Colorado	106	48	58
GIH	Gujarati Indians in Houston, Texas	97	56	41
JPT	Japanese in Tokyo, Japan	113	57	56
LWK	Luhya in Webuye, Kenya	99	45	54
MEX	Mexican ancestry in Los Angeles, California	53	26	27
MKK	Maasai in Kinaywa, Kenya	105	43	62
TSI	Toscani in Italia	102	53	49
YRI	Yoruba in Ibadan, Nigeria	137	68	69
	TOTAL	1096	522	574

Table S2- SNPs removed from the dataset after BLAST filtering. These SNPs were removed from the dataset after performing a sequence similarity search using BLAST of the flanking sequences of autosomal SNPs (±30bp around the SNP) against the Y chromosome sequence. Any SNP with flanking sequences having a perfect match with Y-linked sequences was removed. The sequences were downloaded from UCSC (hg18 assembly).

Chr	SNP code	Flanking sequence -30bp	SNP	Flanking sequence +30bp
1	rs10915767	TCTCTTTTGATCTTTGTTGGTTTAAAGTCT	G/A	TTTTATCAGAGACTAGGATTGCAACCCCTG
1	rs11810671	AGCTTGCCTGGCAGAAGTCTAAGCAAAGGG	G/T	CACAACACTCTTGGCAAGTTGAGAGTGGCA
1	rs12403708	AGAACTTGCCAAATATCCAAGCCCCAAAAA	C/A	CGCACCTGCTGCAATCCCCAAAAGAATAGC
1	rs12743401	AAGAAAAAATTGTCCATAACCATCCATAAC	T/C	ATACTTTTCAGAAGAAATTATTAACATTCT
1	rs4658379	ATCATGTCCTTTGCAAGAATATAAATGGAG	C/A	TGCAGGCCATTATCCTTAGCAAACTAATGC
1	rs7364930	CGGGAGCGAACTACCAACTTTGGGAGGTGA	A/C	ATATCTTACAGAAGGAGTGTGGGAGAAGAA
1	rs7533316	CCACACTGGGACCTTCTGGCTCCTGGATGT	G/A	ACTAGGTTCTACTTCAGGGTCTCCACATAT
1	rs7549760	ATCACAACTCTTCATAACTCTTGGTTTGGC	A/G	GGATACAACCTATCACTGATGCAGTGGTAC
1	rs805909	CCTAGGCCTTCTTTTCCTCCTCATTGAATT	T/C	ACTGACCATTTCTTCTATTGAATTTAAGTA
2	rs10874458	TATAACAGTCATAAGACTGTCATTCTTAGA	A/C	AGGCCTACATGCAAAACTGGCCCTTTGCTG
4	rs4962914	CAATTAAACAACAAAAAGAAATAAGAAAAA	A/G	TTAGCTGGGCATGGTAGTATGCATCTGTAG
4	rs6448544	GGAGCACCCAGATTCATAAAGCAAGTCCTG	C/A	GTGACCTACAAAGAGACAATGTCTTATTTT
4	rs9761692	TGATTGAATTTATTACTTCACTTAACAAAT	A/G	TTCATTACATGCCCACTGTTTGTCAGATAT
5	rs36019094	ATAATCTTCTGATTGGATTGATCAAGAAAG	C/A	CATTAAACTTGGTTGAAGAAGGTAGTCAGG
6	rs9346294	AAACTCTCAATAAATTAGGTATTGATGGGA	C/T	GTATCTCAAAATAATAAGAGCTATTTATGA
7	rs10236123	AACAGATGCTGGAGAGGATGTGGAGAAATA	G/A	GAATGCTTTTACACTGTTGGTGGGAGTGTA
8	rs11786942	CTCAGCTTTGTTCTTTTGGCTTAGGATTGT	C/G	TTGGCAATGCAGGCTCTTTTTTGGTTCCAT
9	rs12551092	CATACATGCCCTGCTCTTGTTTACACTGCC	A/G	GTTTACACTGTTTCTCCAAGCCATCACAGC
9	rs3121249	GATGACCTCATGGGGACTTGCCTTTGAGCA	G/A	TTGACACAGGAAGGGAGGACTAGGGCCTGG
10	rs10794037	AGTAGGATCTGTGACTTGTTTCTAACTGAC	G/A	TAATATTGCAAATGTGATGGGTTGTTACTC
11	rs12271307	AGAAGATGGCTTCCAATAGAAACCAGTTGC	C/T	TCAAAACTGAAAACTAAGTTTGCAGACAGC
11	rs7294016	ATCCATGAGCATGGAATGTTTTTCCATTTG	G/C	TTGTGTCCTCTCTTATTTCCTTGAGCAGTG
12	rs12425187	GTTCTTCCATTTGTTTGTGTCCTCTTTTAT	T/G	TCATTGAGCAGTGGTTTGTAGTTCTCCTTG
12	rs9706434	TTTACAGACAAGCAAATGCTGAGAGATTTT	G/A	TCACCACCAGGCCTGCCTTACAAGAGCTCC
15	rs11638630	TAAAATATCCATCACCCCAGGAAACTTAAC	C/T	TTGAGCACAAACTCTACAACATGTTCAATG
15	rs11857958	AAACAAGCAATGGGGAAAGGATTCCCTATT	T/G	AATAAATGGTGCTGGGAAAACTGGCTAGCT
15	rs150963	GGGGTTTGAAATAGACTCATACTGAATTCT	G/A	TGCTATCGTGGAGTCCATGGTGGTATGAAC
15	rs17861107	GCCACTGCACTCCAGCCTGGGTGACAGAGC	G/A	AGACTCCATCTCAAATAAAAGACTGCAACT
15	rs4486824	AATTCACACATAACAATATTAACCTTAAAT	G/A	TAAATGGGCTAAATGCTCCATTGTTGGTTT
16	rs391728	AGTTGAACCTTTCTTTTGATTGAGCAGTTT	T/G	GAAACACTCTTTTTGTAGAATCTGCAAGTG
18	rs11661664	TCTCTGATGACCAGTGATGATGAGCATTTT	T/C	TCATGTGTCTGTTGGCTGCATAAATGTCTT
18	rs12967815	TAGGACCCTCCGAGCCAGGTGTGGGATATA	G/A	TCTCGTGGTGCGCCGTTTTTTAAGCCGGTC
18	rs4939894	GGGAGCTGTAGACCAGAGCTGTTCCTATTC	G/A	GCCATCTTGGCTCCTCGATTTCCAATTTTA
19	rs35387182	TCAGCCTCCCAAAGTGCTGGGATTACAGGC	A/G	TGAGCCACCACATCTGCTTGCTCCCAATTT
20	rs34165968	AGCTCTTTAGTTTAATTAGATCCCATTTGT	C/A	AATTTTGTCTTTTGTTGCCATTGCTTTTGG
22	rs5993671	TATTTCCTTCTTTGTGTTCATAAGTCCTTA	G/T	GATTTAGCTACCACTGATAGGTGAGAACAT

Table S3 – Selective model used in Fry (2010) for an autosomal locus and an X-linked locus under SA selection. h and s represent respectively the dominance and the selective coefficients in males (h_m, s_m) and females (h_f, s_f) and are comprised between 0 and 1. For the X-linked case, the genotypes are indicated for males-females.

Autosomal case			
Genotypes	**AA**	**AB**	**BB**
Male fitness	1	$1-h_m s_m$	$1-s_m$
Female fitness	$1-s_f$	$1-h_f s_f$	1
X-linked case			
Genotypes	**AY-AA**	**∅-AB**	**BY-BB**
Male fitness	1		$1-s_m$
Female fitness	$1-s_f$	$1-h_f s_f$	1

Table S4- Number of genes and SNPs that are sexually differentiated in a various number of populations (from 1 to 4 populations).

Number of populations	1	2	3	4	total
genes	4988	672	43	2	5705
SNPs	22754	812	11	0	23577

Table S5. P-values of the observed genetic distances between males and females computed on the autosomes for the 11 HapMap populations.

population	Autosomal p-values
ASW	0.84
CEU	0.77
CHB	0.61
CHD	0.35
GIH	0.57
JPT	0.41
LWK	**0.03**
MEX	0.78
MKK	0.82
TSI	0.39
YRI	0.77

Table S6. The p-value of the observed genetic distances between males and females on the X chromosome for the first subsampling of females (over 30) and the number of repetitions with a p-value < 5%.

population	p-value (first repetition)	Number of repetition with a p-value < 5%
ASW	1	0
CEU	0.35	0
CHB	**0.005**	30
CHD	0.26	1
GIH	0.99	0
JPT	0.78	0
LWK	0.95	0
MEX	0.45	0
MKK	0.99	0
TSI	0.98	0
YRI	1.0	0

Table S7- List of genome-wide SD SNPs and comparison of their allelic frequencies in the HapMap 3.3 and the 1000 Genome (phase 1) datasets. The physical positions are taken from the human genome build NCBI36/hg18. The p-values of Fisher's exact tests comparing allelic frequencies between males and females are reported after FDR correction at the genome-wide scale (q-value). Frequencies of the reference allele (allele 0) are given in males and females for the two datasets. The corresponding numbers of chromosomes are given in parenthesis. Inconsistent allelic frequencies are indicated in grey. SNPs in grey boxes were removed. For the CHD population, we compared the HapMap allelic frequencies with the 1000 Genome frequencies in CHB. SNPs with 'suspected' status in dbSNP were removed and are not shown in the table. For rs17315996 (in MKK), we found that CHB showed a high F_{ST} between the sexes as well and compared the allelic frequencies between the two datasets for CHB. We found that they were not concordant and therefore removed this SNP from further analysis.

				HapMap				1000 Genomes	
Chrom	SNP	position	Pop	F_{ST}	q-value	Freq 0 Male	Freq 0 Female	Freq 0 Male	Freq 0 Female
X	rs12394407	51868390	CHB	0,39	2,2E-03	0.33 (51)	0.02 (164)	0.00 (44)	0.00 (106)
			JPT	0,38	1,4E-03	0.33 (55)	0.01 (110)	0.00 (50)	0.00 (78)
X	rs16980138	15711203	CHB	0,43	1,5E-04	0.73 (51)	1.00 (168)	1.00 (44)	1.00 (106)
			JPT	0,40	4,7E-04	0.69 (55)	1.00 (112)	1.00 (50)	1.00 (78)
X	rs10429791	10620289	CHD	0,41	3,1E-03	0.29 (45)	0.00 (116)	0.00 (44)*	0.01 (106)*
X	rs2961384	40217230	CHD	0,41	3,1E-03	0.67 (46)	0.99 (116)	0.95 (44)*	0.99 (106)*
X	rs5973231	34636210	CHD	0,71	4,2E-14	0.40 (48)	1.00 (116)	1.00 (44)*	1.00 (106)*
X	rs5973480	35717997	CHD	0,71	5,9E-13	0.59 (44)	0.00 (116)	0.00 (44)*	0.00 (106)*
X	rs5943145	110508468	GIH	0,63	6,4E-11	0.59 (54)	0.00 (78)	NA	NA
X	rs7053144	111322009	GIH	0,64	2,6E-11	0.39 (56)	1.00 (78)	NA	NA
X	rs17315996	41631945	MKK	0,44	8,6E-03	0.64 (42)	0.98 (122)	NA	NA
3	rs6442608	16563277	YRI	0,20	1,4E-02	0.47 (136)	0.80 (138)	0.76 (86)	0.80 (90)

*Data extracted from CHB population

Table S8- Enriched terms and their p-values in each functional gene cluster for SDG_{3000}, SDG_{1000} and SDG_{sh}. Details of the functional enrichment analysis results from the clustering tool performed on DAVID for the 3000 and 1000 SD genes with the lowest p-values (SDG_{3000} and SDG_{1000}) and the SD genes shared by at least two populations (SDG_{sh}).

	SDG_{3000}	SDG_{1000}	SDG_{sh}
Enriched term	**p-value**		
EPIDERMIS DEVELOPMENT			
Enrichment Score	2.51	2.65	
Cornified envelope	1.56E-02	1.26E-02	
Ectoderm development	2.49E-03	3.21E-03	
Epidermal cell differentiation	2.57E-03	6.39E-04	
Epidermis development	3.66E-03	1.40E-03	
Epithelial cell differentiation	6.11E-04	3.90E-03	
Epithelium development	3.40E-02	4.50E-02	
Keratinization	4.71E-04	4.50E-02	
Keratinocyte differentiation	2.38E-03	3.13E-04	
RESPONSE TO STIMULUS AND IMMUNE SYSTEM			
Enrichment Score	1.38	1.37	
Positive regulation of defense response	1.15E-01		
Positive regulation of inflammatory response	1.62E-01		
Positive regulation of response to external stimulus	1.66E-03	4.28E-03	
Positive regulation of response to stimulus	6.24E-03		
Regulation of inflammatory response	4.76E-01		
Regulation of response to external stimulus	5.66E-02	6.04E-02	
Positive regulation of behavior		4.81E-02	
Positive regulation of chemotaxis		2.88E-02	
Positive regulation of locomotion		4.63E-02	
Positive regulation of positive chemotaxis		4.01E-02	
Regulation of behavior		1.16E-01	
Regulation of chemotaxis		3.59E-02	
Regulation of locomotion		1.61E-01	
Regulation of positive chemotaxis		4.01E-02	
BINDING TO NUCLEOSIDE, NUCLEOTIDE AND RIBO-NUCLEOTIDE			
Enrichment Score	1.59		
Adenyl nucleotide binding	1.76E-02		
Adenyl ribonucleotide binding	1.80E-02		
ATP binding	2.87E-02		
Nucleoside binding	3.35E-02		
Nucleotide binding	1.22E-02		
Purine nucleoside binding	3.08E-02		
Purine nucleotide binding	3.26E-02		
Purine ribonucleotide binding	3.65E-02		
Ribonucleotide binding	3.65E-02		

GLYCOLYSIS			
Enrichment Score	1.43		
Carbohydrate kinase activity	1.89E-03		
Glycolysis	2.24E-01		
Hexokinase activity	1.22E-01		
REPRODUCTIVE PROCESS			
Enrichment Score		1.68	
Gamete generation		1.11E-02	
Male gamete generation		4.80E-02	
Multicellular organism reproduction		1.64E-02	
Reproductive process in a multicellular organism		1.64E-02	
Sexual reproduction		1.24E-02	
Spermatogenesis		4.80E-02	
FILAMENT CYTOSKELETON			
Enrichment Score		1.33	
Intermediate filament		8.74E-02	
Intermediate filament cytoskeleton		9.86E-02	
Keratin filament		1.18E-02	
VACUOLE			
Enrichment Score			1.55
Vacuolar membrane			5.85E-03
Vacuolar part			1.11E-02
Tubulin binding			7.59E-02
Vacuole			1.25E-01
ODONTOGENESIS			
Enrichment Score			1.48
Odontogenesis of dentine-containing tooth			5.72E-03
Odontogenesis			2.19E-02
Tissue morphogenesis			2.90E-01
MICROTUBULE AND CYTOSKELETON			
Enrichment Score			1.46
Microtubule			7.21E-03
Microtubule cytoskeleton organization			1.24E-02
Microtubule-based process			1.92E-02
Microtubule cytoskeleton			8.81E-02
Cytoskeleton organization			3.15E-01
PEPTIDASE INHIBITOR ACTIVITY			
Enrichment Score			1.42
Endopeptidase inhibitor activity			2.96E-02
Enzyme inhibitor activity			3.16E-02
Peptidase inhibitor activity			3.97E-02
Serine-type endopeptidase inhibitor activity			5.67E-02

Table S9- Number of individuals being part of a couple and having at least one child in the HapMap populations. These individuals are part of the HapMap trio dataset.

Population		Number of individuals: total	Number of individuals: adults in trios	proportion
ASW	African ancestry in Southwest USA	41	28	0.68
CEU	Utah residents with Nothern and western Europe ancestry	106	104	0.98
MEX	Mexican ancestry in Los Angeles. California	53	48	0.91
MKK	Maasai in Kinaywa. Kenya	105	43	0.41
YRI	Yoruba in Ibadan. Nigeria	137	109	0.80